DOCUMENT RESUME

ED 128 377 TM 005 499

AUTHOR	Garwick, Geoffrey
TITLE	Program Evaluation Project Report, 1969-1973. Chapter Five: A Construct Validity Overview of Goal Attainment Scaling.
INSTITUTION	Program Evaluation Resource Center, Minneapolis, Minn.
SPONS AGENCY	National Inst. of Mental Health (DHEW), Rockville, Md. Div. of Mental Health Services Program.
PUB DATE	Jun 74
GRANT	NIMH-5-R01-1678904
NOTE	21p.; For related documents, see TM 005 495-501
AVAILABLE FROM	Program Evaluation Project, 501 Park Ave. South, Minneapolis, Minnesota 55415 ($1.00)
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Evaluation Methods; *Goal Orientation; *Measurement Techniques; *Mental Health Programs; *Program Evaluation; Scores; *Validity
IDENTIFIERS	*Goal Attainment Scaling

ABSTRACT

The P.R.P. Report 1969-1973 focuses on the various findings and activities of the Program Evaluation Project. The establishment of validity is one of the major tasks of the developers of a measurement methodology. In this chapter, it is argued that the construct validity approach is essential to an understanding of the validity of Goal Attainment Scaling, since there are no clear criteria available for concurrent validation. The chapter discusses data from a variety of studies on Goal Attainment Scaling in an effort to illustrate various facets of the construct validity approach when applied to the methodology. The findings underscore the idea that Goal Attainment Scaling can be applied in a variety of settings. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED128377

Chapter Five:

A Construct Validity Overview of Goal Attainment Scaling

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

P.E.P. REPORT 1969 - 1973

TM005 490

A Report on Four Years of Staff Effort at the Program Evaluation Project.

CHAPTER FIVE

Program Evaluation Project Report, 1969-1973

A CONSTRUCT VALIDITY OVERVIEW OF GOAL ATTAINMENT SCALING

Geoffrey Garwick

June, 1974

Thomas J. Kiresuk, Ph.D., Director
Program Evaluation Project
501 Park Avenue South
Minneapolis, Minnesota 55415

Developed under Grant #5 R01 1678504, National Institute of Mental Health, Department of Health, Education, and Welfare

---

Acknowlegements are given to all of the Program Evaluation Project Staff, especially Donna Audette, James Baxter, Joan Dreyer, Susan Jones and Laurence Kivens for his valuable editorial comments. Dr. Robert Sherman gave helpful advice and Dr. David Weiss contributed subtantial improvements to the chapter.

TABLE OF CONTENTS

	Page Number
General Introduction to the P.E.P. Report 1969-1973	1
Synposis	2
Introduction	3
I. Variables Relevant to Goal Attainment Scaling Construct Validity	5
II. Generating and Representing Hypotheses about Goal Attainment Scaling	7
III. Validity Measures for Goal Attainment Scaling	8
Conclusion	14
Program Evaluation Project Staff Listing	16

For further information, please contact Ms. Joan Brintnall, Program Evaluation Project, 501 Park Avenue South, Minneapolis, Minnesota 55414.

GENERAL INTRODUCTION TO THE P.E.P. REPORT 1969-1973

The P.E.P. Report 1969-1973 focuses on the various findings and activities of the Program Evaluation Project. It is being published in pamphlet form with one pamphlet for each chapter.

As of January, 1974, the Program Evaluation Project is funded by a three year collaborative grant with the Mental Health Services Division of the National Institute of Mental Health. The purpose of the grant is to emphasize the coordination and dissemination of information on a variety of program evaluation methodologies. Currently, it is expected that the title of the organization will be changed to the Program Evaluation Resource Center during 1974.

Further information on the Goal Attainment Scaling methodology and program evaluation is available in other written and recorded materials from the Program Evaluation Project office. Chapter One, "Basic Goal Attainment Scaling Procedures", Chapter Three, "An Introduction to Reliability and the Goal Attainment Scaling Methodology", and Chapter Nine, "Evaluation of the Adult Outpatient Program, Hennepin County Mental Health Service" of the P.E.P. Report 1969-1973 are now available. Additional chapters will be released this year as they are completed.

SYNOPSIS FOR CHAPTER FIVE

A CONSTRUCT VALIDITY OVERVIEW OF GOAL ATTAINMENT SCALING

PURPOSE: The establishment of validity is one of the major tasks of the developers of a measurement methodology. In this chapter, it is argued that the construct validity approach is essential to an understanding of the validity of Goal Attainment Scaling, since there are no clear criteria available for concurrent validation.

The chapter discusses data from a variety of studies on Goal Attainment Scaling in an effort to illustrate various facets of the construct validity approach when applied to the methodology. The following chapter presents the results of one particular study of the validity of Goal Attainment Scaling.

MAJOR FINDINGS: It is emphasized here that there have been many approaches to validity, but the Cronbach-Meehl concept of construct validity seems to be the most inclusive. If there is a basic construct underlying Goal Attainment Scaling it is the "attainment of expectations", but this major construct is accompanied by many other variables related to the many possible ways in which Goal Attainment Scaling can be applied.

A list of those accompanying variables and a system for illustrating hypotheses, about Goal Attainment Scaling are presented. This system is utilized to present several findings, the clearest being that as predicted, the Goal Attainment score is not significantly related to client characteristics such as age, sex, education, marital status or intelligence. In one study of adult outpatients and day treatment cases, Goal Attainment scores based on therapist scoring were correlated from a .58 to .84 with two questions of global ratings of treatment outcomes answered by the therapists. The correlations of the Goal Attainment score with the consumer satisfaction Index was .23 for one group with correlations for individual consumer satisfaction items ranging from -.12 to .46. The Goal Attainment score was shown to be correlated .31 with predictive accuracy for one group of adult outpatients. In general, correlations with other measures, as expected, are positive, but low to moderate, with early results from the drug effectiveness study, for example, where all concurrent validity coefficients were .52 or less.

In general, the original Kiresuk-Sherman hypotheses about Goal Attainment Scaling are supported. The findings underscore the idea that Goal Attainment Scaling can be applied in a variety of settings.

The two central concepts in the analysis of measurement systems have historically been validity and reliability. The reliability of Goal Attainment Scaling is discussed at some length in Chapters Three and Four of the P.E.P. Report 1969-1973.

This chapter presents suggestions for approaching the study of the "validity" of Goal Attainment Scaling. The Introduction is an attempt to link traditional psychometric comments on validity, especially construct validity, to the special characteristics of Goal Attainment Scaling. Section I is an outline of the variables available for validity studies of Goal Attainment Scaling by the Program Evaluation Project staff. Section II discusses briefly the generation of hypotheses related to Goal Attainment Scaling. Section III shows Program Evaluation Project data applied to issues in Goal Attainment Scaling validity.

Introduction

As Ebel (1961) comments, validity is an imprecise term due to "logical and operational limitations of the concepts of validity itself". He argues that "...faster progress will be made toward more educational and psychological tests if validity is given a much more specific and restricted definition than is usually the case, and if it is no longer regarded as the supremely important quality of every mental test". The classic statement on validity is Lindquist's (1942) observation that "the validity of a test may be defined as the accuracy with which it measures that which it is intended to measure or as the degree to which it approaches infallibility in measuring what it purports to measure". Loevinger (1957) argues that "this definition is too vague, too remote from factual measuring operations, to be useful ..." Ebel also argues that this definition is too general and not particularly useful. He proposes that these points of test "quality" be examined instead:

1. The importance of the inferences that can be made from the test scores.
2. The meaningfulness of the test scores, based on:
   a. An operational definition of the measurement procedure.
   b. A knowledge of the relevance of the scores to other measures, from,
      i. Validity coefficients, predictive and concurrent.
      ii. Other correlation coefficients or measures of relationship.
   c. A good estimate of the reliability of the scores.
   d. Appropriate norms of examinee performance.
3. The convenience of the test in use.

Ebel concludes that these concepts answer the primary questions about the utility of a test, and that points 2a, 2b-i, and 2b-ii need not be established for all tests.

A more stringent approach is presented by Raymond Cattell (1964) who argues that some very mathematically inclined "psychometricians have sometimes seemed lost in their labyrinthine fastnesses from logic, from common sense, and certainly from psychological perspective". He suggests that validity is "...the capacity of a test to predict some specified behavioral measure (or set of measures) other than itself". This use of "predict" could be taken, for the purposes of outcome evaluative measures, in the very broadest sense, such as "a high score for treatment X on measure Y for group Q predicts a high outcome for other groups similar to group Q who receive treatment X". In any case, Cattell specifies three parameters of validity:

1. "Degree of Abstraction of the Referent Criterion." This parameter might also be called, the generalizability of results, with one extreme being a very concrete test, such as "ability to clean armadillos quickly", and the other extreme being a very conceptual test, such as "creativity".
2. "Degree of Naturalness of the Criterion." This parameter varies from a criterion based on a "Natural" situation such as coping with a real disaster, to correlation with an "Artificial" criterion situation, such as another test.
3. "Degree of Directness of Validation." This parameter varies from "direct" or complete correlation with the criterion to "indirect or circumstantial" correlation with the criterion. The directness refers to the patterns by which variables are correlated. For example, two measures could both be correlated .3 with criterion X, and yet measure #1 could correlate with only one aspect of criterion X, whereas measure #2 could relate to an entirely different aspect of criterion X.

After introducing these three parameters, Cattell discusses some of the concepts frequently called "validity". He maintains that there are a group of "utility coefficients" which are not validity coefficients at all and includes "face validity", "content validity" and "semantic validity" in this group of non-validity concepts. Mosier (1947), writing much earlier, also recommends that "face validity" be dropped from the vocabulary because it has too many

confusing and vague definitions. Cattell (1964) appears to see the "...integration of psychometrics with personality theory and general psychological theory..." as an ultimate goal for improving validity conceptualizations, in much the same way that Loevinger (1957) stresses it. In fact, Loevinger maintains that "content validity is established by the judgment of the investigator that the items are valid; it is thus also contingent upon a special, non-generalizable circumstance, to wit, the particular investigator.... Since ad hoc arguments are scientifically of minor importance, if not actually inadmissable, what is left, construct validity, is the whole of the subject of a systematic scientific point of view".

The viewpoints presented above reveal the range in theoretical thinking about validity. In this discussion, construct validity is utilized as a basis for the presentation, since it allows for flexibility and subsumes many earlier "forms" of validity.

This approach to the validity issue is based on Cronbach and Meehl's (1955) version of the concept of "construct validity", which they added to earlier "types" of validity: such as concurrent, content, and predictive. It is very germane to Goal Attainment Scaling theory that Cronbach and Meehl observe (evidently citing Gaylord), "When an investigator believes that no criterion available to him is fully valid, he perforce becomes interested in construct validity because this is the only way to avoid the infinite frustration of relating every criterion to some more ultimate standard... Construct validity must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured."

This lack of a clear-cut criterion for comparison is a central issue in the discussion of Goal Attainment Scaling validation. There is no clear-cut criterion for either mental health or therapy effectiveness in mental health. With the use of Goal Attainment Scaling, in effect, a new criterion for mental health treatment is selected, that is, "the degree to which expectations for treatment are achieved." The expectations may be developed by the clinicians, the clients, other persons, or some combination of involved persons. Since this form of criterion is so new, no immediately applicable standard for comparison with Goal Attainment Scaling has been located. Although some instruments may be suitable for examining certain aspects of Goal Attainment Scaling, none have been identified which are as comprehensive or "expectation-oriented" as Goal Attainment Scaling.

If a clinician developed the expectations as appearing on a particular follow-up guide, the result could be taken theoretically as a one person sample of the expectations for treatment which would be imposed by the hypothetical average clinician. In any case, as noted above, the use of individualized expectations as a criterion is so unusual that there are no easily applicable concurrent measures and a construct validity approach appears essential.

The construct, for Meehl and Cronbach, is "...some postulated attribute of people assumed to be reflected in test performance.... A construct has certain associated meanings carried in statements of this general character: Persons who possess this attribute will, in situation X, act in manner Y (with a stated probability)." It is noteworthy that, even in this very pertinent article, the language of psychometrics is not well-adapted to outcome evaluation in general, or Goal Attainment Scaling in particular. For purposes of the discussion of Goal Attainment Scaling, the above phrase would have to be changed to read, perhaps, "Agencies, treatment modes, or persons who possess this attribute will, in situation X, act in manner Y or will have acted in manner Y (with a stated probability)."

Despite their relatively straightforward definition of a "construct" the authors emphasize that there is no single, simple coefficient of construct validity, but that construct validity must be thought of in terms of a "nomological net" of constructs linked by testable hypotheses. Meehl and Cronbach (1955) delineate construct validity by a series of axioms:

1. Constructs of varying degrees of definitiveness are defined by a network of propositions.
2. There must be predicted relationships among variables.
3. The network must be explicit.
4. Many types of evidence are relevant to construct validity and both high and low correlations may be useful evidence for the proposed nomological net.
5. When a predicted relationship fails to be observed, the network of constructs must be redefined.
6. There is no simple coefficient of construct validity.
7. General scientific procedures are used.

Cronbach and Meehl's discussion, like that of most of the authors mentioned before, is based largely on the concept of the trait or characteristic. An evaluation technique, however, is not usually aimed at measuring a trait of a person, but rather at measuring a pattern of changes or effects in an agency (which may be measured either for the agency as a whole or through the sum of effects on a number of persons). Wiggins (1973) describes a "trait" as "...a hypothetical construct which provides an organizing principle for relating a variety of superficially dissimilar behaviors under a single dispositional unit". If there is a trait underlying the Goal Attainment score, it would be "the tendency to attain expectations" as noted above. The Goal

Attainment expectations could be set by one more of a number of sources, such as the client, the therapist, and so on. Clearly, even if Goal Attainment Scaling can be linked to trait concepts, the methodology demands a loose application of many validity ideas.

These experts on psychometrics have been cited above to suggest the flexibility of the "validity" concept. Possibly the most basic characteristic of validity is that, as Nunnally (1967) says, "Validity is a matter of degree rather than an all-or-none property, and validation is an unending process.... Strictly speaking, one validates not a measuring instrument, but rather some use to which the instrument is put." Nunnally concludes that, "... sufficient evidence for construct validity is that the supposed measures on the instrument ... behave as expected".

This general commentary on validity has been used to introduce a few construct-oriented guidelines. Possible variables related to the "attainment of expectations" construct underlying Goal Attainment Scaling are outlined in the first two sections. Then in the third section, findings on validity will be summarized for Goal Attainment Scaling.

I. Variables Relevant to Goal Attainment Scaling Construct Validity

As observed above, Goal Attainment Scaling is not a typical "trait-type" problem in validity. Instead, the methodology involves a collection of characteristics and activities which can be interconnected by predictions and hypotheses. It is the utility of Goal Attainment Scaling in evaluation methodology which must be examined in the contest of a network of constructs.

Ebel's three points related to an instrument's "quality" or validity were presented in the first page of this discussion. His divisions of "quality" seem appropriate to the considerations of the practical issues of validating an outcome-oriented methodology like Goal Attainment Scaling. Thus, in the third part of this chapter where construct validity issues are empirically examined, Ebel's outline will be followed as a matter of convenience.

This presentation is not intended to be a contribution to the reliability versus validity controversy. Ebel (1961), as mentioned previously, includes reliability as a basic factor in test "quality". Nunnally (1967) states that " ... consistency is a necessary but not sufficient condition for construct validity". Therefore, reliability is subsumed in part of the following discussion, although a more thorough description of these reliability results is presented in Chapters Three and Four of the P.E.P. Report 1969-1973.

In summary, the validity issues for this discussion of Goal Attainment Scaling are oriented to a construct validity perspective in which basic questions are asked about the methodology's utility and measurement properties in the program evaluation context. (The effectiveness of utilizing Goal Attainment Scaling as part of the therapeutic process is not within the scope of this chapter, except in reference to specific measurement studies.) This perspective may include correlations of the Goal Attainment scores with criteria, a procedure often called "concurrent validity", or content analyses which could be related to the so-called "content validity", but the basic thrust of the discussion is to examine the network of relationships between constructs, the Goal Attainment scores and other variables. Anastasi (1970) comments bluntly, however, that "...content, criterion-oriented, and construct validity do not correspond to distinct or logically coordinate categories. On the contrary, construct validity is a comprehensive concept, which includes the other types."

The primary construct being examined is "outcome" or "attainment of expectations". "Outcome" could actually be considered a collection of different constructs. For example, "outcome after one month of treatment" is certainly a different set of expectations, and a different construct, with different postulated attributes, different predictions, and a different meaning than outcome at six months."

The basic components of Goal Attainment Scaling procedures may be considered in terms of characteristics and activities. Characteristics are attributes of the persons involved in the Goal Attainment Scaling situations and activities are considered to be behaviors or procedures of the Goal Attainment Scaling process. Both aspects are represented here by the data available from Program Evaluation Project data, some of which is incomplete.

Characteristics

A1. Of the persons (or agency) being represented on the Goal Attainment Follow-up Guide.

A2. Of the person(s) constructing the Goal Attainment Follow-up Guide.

A3. Of the person(s) scoring the Goal Attainment Follow-up Guide.

Activities

B1. The rules and procedures used to construct the Goal Attainment Follow-up Guide.

B2. The treatment being used in the agency.

B3. The way in which the Goal Attainment results are expressed.

B4. The way the Goal Attainment Scaling results are used.

Here are some more specific variables included under the seven major components mentioned above:

A1. The person represented on the follow-up guide. (If organizations were being represented on the follow-up guide, their characteristics would be listed here, such as size, income, available evaluative resources and so on.)

a. Truthfulness, completeness and ability to communicate and/or predict.

b. Age, sex, intelligence, and other variables.

c. History of treatment.

d. Problems presented and diagnosis.

A2. The person constructing the Goal Attainment Follow-up Guide.

a. Who is it? the client, a spouse of the client, a relative, a clinician(s), some combination of the above persons, or others?

b. How experienced is the constructor at Goal Attainment Scaling and has the constructor been trained in Goal Attainment Scaling?

c. How accurate is the constructor's ability to predict outcome?

d. Personality and demographic characteristics of the constructor and/or educational or discipline background.

A3. The person scoring the Goal Attainment Follow-up Guide.

a. Who is it? the client, the person constructing the follow-up guide, the person giving the treatment or a separate person?

b. How experienced and skilled is the person at scoring the Goal Attainment Follow-up Guide?

c. Personality, educational, demographic and discipline characteristics of the scorer. (If other than the client?)

d. How skilled at interviewing is the person scoring?

B1. Rules and procedures used to construct the Goal Attainment Follow-up Guide.

a. What form of the Goal Attainment Follow-up Guide is used?

b. What rules of construction are used and are the follow-up guides rechecked to see if the rules are met?

c. Is the Goal Attainment Follow-up Guide aimed at a specific date in the future and how far off is that date?

d. Is the level at intake used?

e. Can scales be differently weighted?

f. Is the Goal Attainment Scaling semi-standardized, standardized or completely idiosyncratic?

g. Is the treatment to be received and/or the person who will be treating known to the follow-up guide constructor?

B2. The treatment being used.

a. What form of treatment is being used and what is available?

b. What rules of treatment choice (e.g. random) are used and is their use monitored?

c. Is the treatment limited in time and how long does it last?

d. Is the Goal Attainment Follow-up Guide available to the treater or used as part part of treatment?

e. Can treatments be changed during the client/treater interaction and how often are they changed in practice?

f. Is there any limit on the type or number of problems to be treated?

B3. The type of measure used to express the Goal Attainment results.

a. Goal Attainment score mean (depends on follow-up results).

i Kiresuk-Sherman Goal Attainment score (varies from 20 to 80).

ii. Scale-by-scale Goal Attainment score, (varies from -2 to +2) for either an individual scale or the mean for an entire follow-up guide.

b. Goal Attainment score variance (depends on follow-up results).

c. Change score (depends on follow-up results and whether the initial status of the client is noted on the follow-up guide).

d. Predictive accuracy also called Mean Inaccuracy Score (depends on follow-up results).

e. Contents or types of problems included on the Goal Attainment Follow-up Guide.

f. Reaching arbitrarily established levels, regardless of expectations or change.

g. Are rules and procedures established for the follow-up scorer?

h. How often are scales unscoreable and how confident is the follow-up scorer in the accuracy of his score?

B4. The way the Goal Attainment score results are used.

a. Who receives the results, supervisors, no one, clients, treaters, legislators?

b. What are the consequences of the results, i.e., information, salary, employment, advancement, publicity, peer pressure, etc.?

c. How long will the results be received and how many times or how often?

d. Will the results be used in conjunction with different measures, i.e., costs?

This array of variables illustrates the immensity of the validation task for Goal Attainment Scaling, and this list is not necessarily complete. There are thirty-seven components listed here, some of which contain multiple variables.

In addition to the three characteristics and four activities listed previously, a related characteristic which should be discussed here is the effect of the experimental design in which Goal Attainment Scaling is being applied. It should be understood that Goal Attainment Scaling is a measurement tool which can be utilized within a range of desired degrees of formal scientific or experimental procedures, such as control groups, random assignment, impartial interviewers, etc. Such experimental procedures are part of the total matrix of the utilization of Goal Attainment Scaling but are not necessarily included in the list of Goal Attainment Scaling activities and characteristics.

Other variables are available for some purposes as partial comparative criteria. These include a variety of commonly utilized outcome correlates and are other methods for measuring what is usually assumed to be some segments of "outcome". Goal Attainment Scaling is designed to be more comprehensive and more sensitive to client outcome than such measures (Kiresuk and Sherman, 1968). Available criteria include:

1. Minnesota Multiphasic Personality Inventory (MMPI) and other personality measures.

2. Intelligence Quotient (IQ) results.

3. Brief Psychiatric Rating Scale.

4. Self-Rating Symptom Scale.

5. Consumer satisfaction scores.

6. Therapist ratings of global improvement.

7. Differences among groups receiving different treatments.

8. Taylor Manifest Anxiety Scale.

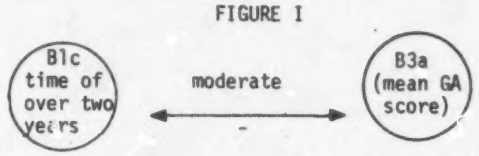
These eight potential criteria, plus the thirty-seven components of Goal Attainment Scaling presented above, total at least forty-five components for the study of Goal Attainment Scaling construct validity. Obviously, many other criteria could be utilized, further expanding the list of components which could be examined.

II. Generating and Representing Hypotheses about Goal Attainment Scaling

Clearly, not all the possible relationships among these activity and characteristic components of Goal Attainment Scaling can be represented, but a cross section of hypothetical links can be suggested for some of the more important components. Predicted relationships among the components will be represented by arrows ($\longleftrightarrow$) with a superscript of "high" for a predicted correlation of over (.70), "moderate" for a predicted correlation of (.40) to (.69), "low" for a predicted correlation of (.39) to (.15) and "no" for a correlation of (-.14) to (.14). Positive correlations are "+" and negative correlations are "-", "Crit" refers to one of the eight criterion variables, but other codes refer to the Goal Attainment Scaling components listed previously.

Construct and variable relationships can be hypothesized and represented for Goal Attainment Scaling theory with this system of activities, characteristics, and criteria. One of the difficulties with the construct validity approach, however, is that standards for the number and nature of the links required in the nomological network have not been clearly established. In addition, since it could be assumed that change is expected over time for "outcome" measures, there are actually different constructs with different sets of variable relationships when there are different lengths of time between follow-up guide construction and follow-up scoring. For example, although age would not be expected to be associated with the outcome construct of Goal Attainment if the follow-up is completed inside a year or two, age could have some effect if the follow-ups were completed only after five years and either children or elderly persons were included as Figure I illustrates.

FIGURE I



These and other theoretical considerations imply that any summary comments about the construct validity and Goal Attainment Scaling should be cautious.

The hypotheses associated with Goal Attainment Scaling are still fairly rudimentary and only a few have been tested extensively. The most firmly established hypothesis is that Goal Attainment scores are relatively unresponsive to the demographic characteristics of either individuals or groups of clients. This lack of relationship was predicted because the very process of setting the Goal Attainment Scaling levels is based on developing expectations which should allow for the unique features, demographic or otherwise, of the client. The evidence in relation to other hypotheses, as discussed below, is somewhat sparser. The collection and analysis of data relevant to the hypotheses continues.

The original Kiresuk and Sherman article of 1968 predicted that the Goal Attainment score should have a low to moderate correlation with already existing outcome measures. Since the Goal Attainment score is 1) based on an individualized measurement system, it should not have a high correlation with non-individualized measures, and 2) since it is specific and goal-oriented, it should differ from global or change-oriented measures even if they are individualized.

Sections of the Hypthetical Nomological Net for Goal Attainment Scaling.

FIGURE IIa:

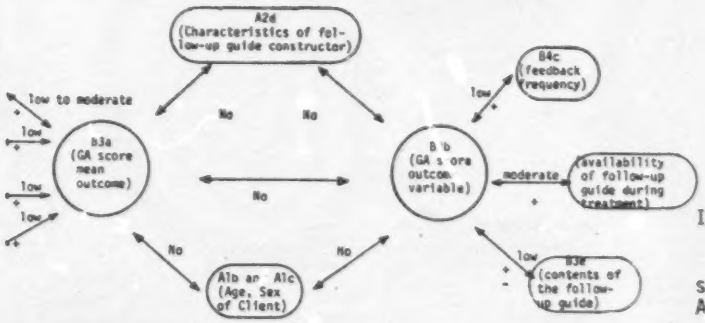


Figure IIa illustrates several hypotheses that could be tested. For example, the part of the diagram in the upper left quadrant of Figure IIa advances the hypothesis that the Goal Attainment score mean, as a representative of the "outcome" construct, (B3a), should have a correlation with criterion 6.

Figure IIb and IIc are also representations of intervariable relationships. Figure IIc suggests that the mean Goal Attainment score should be slightly related to the accuracy of predictions by the follow-up guide constructor and to the relationship of the follow-up guide constructor to treatment.

FIGURE IIb:

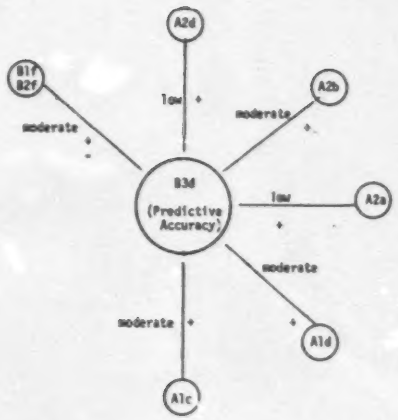
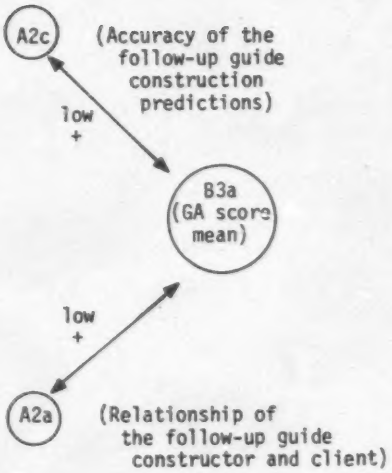


FIGURE IIc:



III. Validity Measures for Goal Attainment Scaling

In the following section, this representational system is used to discuss some hypotheses about Goal Attainment Scaling with pertinent data.

The three major areas of test "quality" established by Ebel will be used to separate the presentation of these components' interrelationships into smaller units for ease of discussion. Ebel's points were presented previously on page 3.

Examples will be presented in relation to each of these points.

A. "The Importance of the Inferences that Can Be Made"

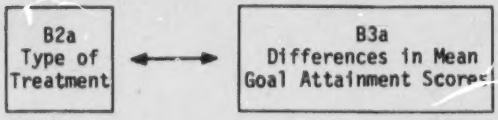
Clearly if a criterion of effectiveness in mental health treatment could be established, it would be a very significant contribution to evaluation and psychopathology knowledge.

Specific examples of investigating treatment effectiveness include:

1. Differences Among Four Outpatient Treatment Modes

In a group of clients randomly distributed among treatment modes, the mean Goal Attainment score should vary among modes which have different degrees of effectiveness.

FIGURE III



The results in the earliest study using Goal Attainment Scaling suggested that differences among modes were not great when all clients were considered as a group regardless of other variables. On June 5, 1972 for 186 nonrandomly assigned subjects, the four modes of therapy in the original Program Evaluation Project study varied only from 48.7 for Day Treatment to 50.2 for Individual Therapy, 50.3 for Drug Clinic to 51.1 for Group Therapy. (The procedures for this study are discussed in other P.E.P. Report 1969-1973 chapters.)

Total results for all 249 randomly assigned clients as of 1973, reveals even less variation among treatment modes in terms of Goal Attainment score.

Treatment Z	50.0
Treatment Y	50.2
Treatment X	50.3
Treatment W	50.8

These means are not statistically different even at the $p < .10$ level and certainly there is little clinical significance to such miniscule differences. (The above data are based on assigned treatment modes, not actual treatment patterns.)

More recent data for the randomly assigned cases in this four-mode study of the Program Evaluation Project, whose procedures are described elsewhere in the P.E.P. Report 1969-1973, are slightly more encouraging, as Table I suggests, when separated by randomization pattern.

There are, however, no Goal Attainment score differences which reach the $p < .10$ level. It is noteworthy that the Consumer Satisfaction Index differences in means also do not reach this level of statistical significance. It is not clear, however, that these treatment would be expected to be significantly different in outcome, especially since the meaning of these various modes were not rigorously defined.

TABLE I: Outcome Scores for Clients Who Stayed in Assigned Treatment Mode at Least Half of Their Treatment Sessions for the Four Mode Study

		TREATMENT W	TREATMENT X	TREATMENT Y	TREATMENT Z
PATTERN I	Mean Goal Attainment Score	47.34 (N=3)	52.28 (N=3)	48.82 (N=3)	47.89 (N=4)
	Mean Consumer Satisfaction Index	57.14 (N=3)	67.77 (N=3)	55.55 (N=3)	81.54 (N=4)
PATTERN II	Mean Goal Attainment Score	47.27 (N=6)	41.41 (N=7)	48.97 (N=9)	
	Mean Consumer Satisfaction Index	63.09 (N=6)	68.97 (N=7)	67.45 (N=9)	
PATTERN III	Mean Goal Attainment Score	52.65 (N=10)	49.04 (N=9)		56.32 (N=14)
	Mean Consumer Satisfaction Index	77.81 (N=10)	76.19 (N=9)		76.36 (N=14)
PATTERN IV	Mean Goal Attainment Score	52.49 (N=119)	52.89 (N=62)		
	Mean Consumer Satisfaction Index	80.29 (N=118)	77.96 (N=61)		

2. Two Treatment Modes at a Day Treatment Center

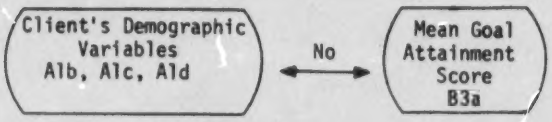
Another study involving randomization of 14 clients between two therapy modes, however, has shown statistically significant differences. In the Day Treatment Center of the Hennepin County Mental Health Service, half of the clients were randomly assigned to a situation in which they prepared their own Goal Attainment Follow-up Guides, while the other half of the clients did not directly set goals for themselves. A clinician constructed standard Goal Attainment Follow-up Guides for both groups. Based on follow-up interviewer ratings of these clinician-prepared follow-up guides, the clients who were involved in their own goal-setting had a mean Goal Attainment score of 71, while the clients who did not set goals had a mean Goal Attainment score of 59, a difference significant at the $p < .015$ level (two-tailed). In addition, clients preparing their own follow-up guides reported greater consumer satisfaction with the significance of the difference in means ranging, in terms of specific questions, from $p < .05$ to $p < .10$.

3. Client Characteristics and the Goal Attainment Score

In constrast to the treatment comparisons made in points number 1 and number 2 above, most non-

treatment variables should not be related to the mean Goal Attainment score. The individualized development of the Goal Attainment Follow-up Guide should take into account such client-specific differences when the expectations on the follow-up guide are set.

FIGURE IV



This no-difference example is significant since it (a) suggests, by contrast, the importance of Goal Attainment score differences due to different degrees of treatment effectiveness and (b) indicates that the Goal Attainment Follow-up Guide can be specially adapted to each client, which is a key assertion of Goal Attainment theory. This portion of Goal Attainment Scaling validation is strongly supported. In January of 1972, a series of linear regressions were calculated for the relationship between the Goal Attainment score of 50.55. (Baxter, Tripp, 1972) The correlations did not reach the $\rho<.10$ level of statistical significance. The highest correlation was for income measures, which were correlated .20 and .18 with the Goal Attainment score. See Table II.

TABLE II: Correlations with Goal Attainment Score Mean (Pearson Product Moment) or Equivalent

AGE	.034
SEX	-.049
MARITAL STATUS	.01
NUMBER OF CHILDREN	.075
"IS INCOME ADEQUATE"	.111
INCOME SOURCE	.184
INCOME AMOUNT	.204
HIGHEST GRADE COMPLETED	.054

Presenting problems mentioned by the clients also had low statistically non-significant correlations with Goal Attainment Scaling according to the data in that study. Some of these presenting problem variables are dichotomous or polychotomous and were correlated accordingly. A few examples are presented here in Table III.

TABLE III: Judgments on a Standardized Information Form Correlated

Severity of Problem (N = 193)	.102
Thoughts of Suicide (N = 178)	.153
"Are Meds being Taken Currently?" (N = 158)	.142

A June, 1972 report by Baxter and Tripp mentioned before suggests that even cross-analysis by age and sex does not lead to statistically different (at the $\rho < .10$ lvel) mean Goal Attainment Scaling scores, as shown in Table IV.

TABLE IV: Goal Attainment Scores by Sex and Age-Group for 186 Clients

Sex	Young	Old
Male	50.11	53.00
Female	50.17	50.57

Thus, in these samples Goal Attainment scores seem to be largely independent of client characteristic variables.

A similar finding is suggested by data released on September 15, 1971 resulting in a low correlation between Shipley-Hartford Intelligence scores and the Goal Attainment scores. (Meade, 1971)

B. "The Meaningfulness of the Test Scores"

This is Ebel's second major aspect of validity. He includes four subpoints, each of which is discussed below.

1. Operational Definition of the Measurement Procedures

A large amount of operational flexibility is basic to Goal Attainment Scaling. This methodology is not a single, rigidly determined set of procedures, but a collection of guidelines from which procedures may be, within some limits, established to fit the needs of each agency. The various Project publications are intended to outline a range of approaches and to describe what has been done in the Project's research. The publications are not designed to establish definitive, permanently fixed procedures, but to allow a range of implementations of Goal Attainment Scaling. (See Programmed Instruction in Goal Attainment

Scaling, Garwick, 1973 and chapter one of the P.E.P. Report, 1969-1973.)

2. Knowledge of the Relationship of the Scores to Other Measures through Coefficients

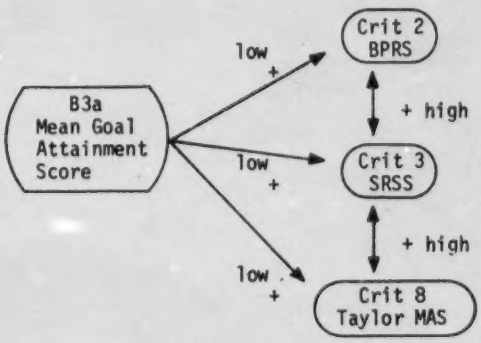
This topic is the second subpoint within Ebel's second general area, "The meaningfulness of the test scores". This aspect of a measurement system would commonly be called "concurrent validity", i.e., validation through other measurement devices. In this instance, the "other measures" used for comparison should be other program evaluation or treatment outcome measures, as opposed to the client characteristic measures discussed earlier.

As suggested earlier, Goal Attainment scores are not intended to have a particularly high correlation with other treatment outcome measurement devices, since Goal Attainment Scaling is such a radically different evaluation system (Kiresuk and Sherman, 1968).

a. Concurrent Validity in the Drug Effectiveness Study of the Campbell-Fiske Matrix

The predicted relationships within some of the major outcome measures of the Program Evaluation Project "Drug Effectiveness Study", appear below in a part of the nomological net.

FIGURE V



The actual results, based on the first block of 20 cases completed for the Drug Study appear in Table V in the following modified Campbell-Fiske (1959) matrix. This matrix is based on data supplied by Baxter and Jones (1973).

This application of the matrix is modified from the Campbell-Fiske concept by the assumption that the clients' outcomes at a given time are the equivalent in the matrix of a "trait". For example, initial status is one trait, outcome at three weeks after intake is another trait, outcome after two months is the fourth trait. There are four variables to measure these traits, except that the Taylor Manifest Anxiety Scale was not administered at at the three week's follow-up. (The places on the matrix which are left blank (because the Manifest Anxiety Scale was not administered at the three week follow-up are marked by a capital "X".) Thus, each of the four measurement points in the Drug Effectiveness Study are utilized in the matrix in the same relationship as a personality trait would be utilized in the original conception of the matrix. The matrix should be interpreted, of course, in terms of the construct relationships expected for Goal Attainment Scaling.

The concurrent or "convergent" validities are represented on the matrix by the numbers not included in the triangles. As predicted, correlations between Goal Attainment Scaling and the other measures are generally low. Except for the correlation between Goal Attainment Scaling and the Manifest Anxiety Scale at the two month follow-up which was .52, the correlations are all below .30, with seven below .20. Only the .52 correlation reaches the $p < .05$ level of significance (two-tailed). By contrast, for the ten concurrent validity correlations not involving Goal Attainment Scaling, (for example, the correlations between the Taylor Manifest Anxiety Scale and the Self-Rating Symptom Scales, which is .73 at the initial measurement, .74 at the two month's follow-up, and .80 at the three month's follow-up) eight are over .30, with five over .50 and three over .70.

The first type of "discriminant validity", which requires that convergent validities be larger than correlations between different measures, in Wiggins' discussion of the Campbell and Fiske Matrix. (Wiggins, 1973) This validity criterion is not met consistently in this study by any of the outcome measurement devices, and particularly not by the Goal Attainment Scaling data, suggesting that correlations between two different outcome measures taken at a single follow-up time are not consistently higher than correlations involving different measures or different follow-up times. Similarly, the second and third types of discriminant validity are not met, suggesting that correlation in outcome due to "method" variance (which in this case is not really "method" variance but rather should be called follow-up variance) is as great or greater than correlation due to different follow-up times.

The solid triangles illustrate the "reliability" correlations. Since in reality, all these outcome measures are based on three different times, some change might be expected in the relative performance of the clients. Out of the six Goal Attainment Scaling correlations, four are below (.20), while for the 15 correlations from the other measures, only two are less than (.30), 11 are greater than (.50) and six are greater than (.60). This pattern of outcome reliability suggests that the Goal Attainment score is less stable than the other outcome measues. Whether or not this smaller degree of stability represents a greater sensitivity to real shifts in outcome will have to be determined by future studies. The Drug Effectiveness Study is continuing and later data may help illuminate the situation.

b. Concurrent Validity and the Consumer Satisfaction Index

Consumer satisfaction results have also been

TABLE V: An Outcome Evaluation Multitrait-Multimethod Matrix for the Drug Effectiveness (Valium versus Psychotherapy) Study

(N=20)

		Method 1 (Goal Attainment Scaling)				Method 2 (Brief Psychiatric Rating Scale)				Method 3 (Self Rating Symptom Scale)				Method 4 (Manifest Anxiety Scale)			
	Traits	Initial	3 Weeks	2 Months	6 Months	Initial	3 Weeks	2 Months	6 Months	Initial	3 Weeks	2 Months	6 Months	Initial	3 Weeks*	2 Months	6 Months
Method 1 (Goal Attainment Scaling)	Initial																
	3 Weeks	-.09															
	2 Months	-.12	.70														
	6 Months	-.22	.20	.47													
Method 2 (Brief Psychiatric Rating Scale)	Initial	.04	-.30	.17	.28												
	3 Weeks	.15	.16	-.20	.08	.11											
	2 Months	-.22	.07	-.26	-.12	.52	.55										
	6 Months	.05	.11	-.23	-.06	-.06	.75	.50									
Method 3 (Self Rating Symptom Scale)	Initial	.21	.11	-.13	.24	.47	.20	.37	.20								
	3 Weeks	.24	.15	-.27	.10	.25	.55	.37	.38	.58							
	2 Months	.12	.28	-.23	.09	.43	.60	.51	.33	.66	.80						
	6 Months	.19	.16	-.39	.18	.04	.51	.20	.42	.47	.52	.64					
Method 4 (Manifest Anxiety Scale)	Initial	-.15	-.15	-.27	.25	.27	.44	.52	.17	.73	.57	.52	.18				
	3 Weeks*	x	x	x	x	x	x	x	x	x	x	x	x	x			
	2 Months	-.30	-.16	-.52	-.33	.27	.28	.38	.05	.84	.70	.74	.42	.84	x		
	6 Months	.33	-.05	-.34	-.07	.11	.51	.24	.20	.61	.69	.75	.80	.49	x	.66	

* The Manifest Anxiety Scale was not scored at the two week follow-up.

compared with the Goal Attainment score. In a February 1974 report, Dreyer noted a Pearson-Product Moment Correlation between Baxter's seven item Consumer Satisfaction Index and Goal Attainment score of .21 for 686 followed-up cases. In a 1973 report, Baxter presented correlations between the Goal Attainment score and 12 of the various individual items on the Consumer Satisfaction inventory for 202 randomly assigned clients. These 12 correlations ranged from -.12 to .46, with four over .20. This report also showed the Goal Attainment score and the Consumer Satisfaction Index which is based on seven items to be correlated .23 (for 199 clients).

c. Concurrent Validity and Therapist Ratings

A useful addition to the Goal Attainment Scaling validation data is the combination Validity/Reliability Study by Baxter. (Baxter, 1973) In this study, therapists were asked to complete three procedures for clients followed up after April 15, 1972: 1) answer two global rating questions about a client's progress in therapy, 2) score the client's Goal Attainment Follow-up Guide before seeing the scores from the follow-up interviewer, and 3) rate the "relevance, "optimism" versus "pessimism", and need for additional scales for each client's scales. The global ratings and the relevance, optimism/pessimism, and need for additional scales are valuable concurrent validation procedures.

It would be expected that the correlations between Goal Attainment scores and the individual global ratings would be somewhat higher than between Goal Attainment scores and more standardized measures such as the BPRS or SRSS as illustrated in Table V. Table VI shows the recorded Pearson Correlations for the Validity/Reliability Study. When therapist global ratings and Goal Attainment scores compare, the correlations range from .582 to .849.

TABLE VI

	"Indicate how well, in your opinion, each patient did in relation to the typical patient in your caseload..."		"Indicate how successful, in your opinion, was your interaction with this patient in relation to the typical patient in your caseload..."	
	Outpatient N = 53	Day Treatment N = 8	Outpatient N = 53	Day Treatment N = 8
Correlation with Therapist Follow-up Score	.582	.689	.604	.849
Correlation with Follow-up Interviewer Follow-up Score	.325	.192	.319	.507

d. Concurrent Validity and the MMPI

Mauger has found a Pearson Correlation of .285 between the Goal Attainment score and the MMPI mean-change Index, and a correlation of .306 between the Goal Attainment change score and the MMPI mean-change score. Neither coefficient reached the $p<.05$ level of statistical significance.

e. Concurrent Validity and Predictive Accuracy

In an analysis of data for the forty-four cases in the original reliability study (see the P.E.P. Report, 1969-1973 chapters on follow-up and reliability) a number of measures were investigated (Twedt, 1974). In this study, for each client, one follow-up guide was constructed by the intake interviewer and a second follow-up guide was constructed by the therapist. Then, these two follow-up guides were combined and this combined follow-up guide was scored independently at two different interviews by two different interviewers.

For the therapist-constructed follow-up guides, the Goal Attainment score was correlated only .18 with the Consumer Satisfaction Index, but was correlated .31 (significant at the $p < .05$ level) with predictive accuracy of the follow-up guide (i.e., mean inaccuracy score, absolute deviation for the deviation for the expected mean of 50). There was no correlation which reached the $p < .05$ level of significance between either the Goal Attainment score for the therapist or for the intake interviewer follow-up guides and variables of age, sex, or number of treatment sessions with the clients, although intake interviewer Goal Attainment score was correlated -.23 with age of the client. For the intake interviewer constructed follow-up guides, the Goal Attainment score was correlated only .24 with the Consumer Satisfaction Index and .16 with predictive accuracy of the follow-up guide. Thus, the predictive accuracy was correlated significantly with the Goal Attainment scores for the therapist but not for the intake interviewer, suggesting that the predictions could have influenced the course of therapy in this case, which is very compatible with the fact that intake interviewers were not involved in the therapy directly.

All these results suggest that Goal Attainment Scaling is not highly correlated with other measurement systems. Nonetheless, there is a positive correlation with the other systems, and the correlations are in the anticipated low to moderate range.

The "Change score" is another possible concurrent measure. Present data suggest moderate correlations between the Goal Attainment score and the Goal Attainment Change score in the .10 to .30 range for various groups. (See the chapter on the Change score in the P.E.P. Report, 1969-1973.)

3. Reliability Measures

Reliability, Ebel's third subpoint in his second general area, "The Meaningfulness of the Test Scored", has been investigated repeatedly for Goal Attainment Scaling. It is not clear that his logic is correct when he included reliability under the "meaningfulness" rubic, but it is included here in order to follow his outline. Extensive discussion of the reliability results are available in the chapter on Reliability and the Goal Attainment Scaling Methodology in P.E.P. Report 1969-1973.

a. One such study, the Interdisciplinary Reliability Follow-up Study, was completed late in 1972. (See the Follow-up Chapter in the P.E.P. Report, 1969-1973.) The study examined differences between interviewer discipline, telephone versus in-person interviewing, and the first versus the second interview. This arrangement of repeated trials will tend to minimize the estimate of reliability, since follow-up interviewers and other variables can change in the interval between the two interviews (a mean of 27 days). (See the chapter on Reliability of Goal Attainment Scaling in the P.E.P. Report, 1969-1973.)

Even in these demanding circumstances, the overall Pearson Correlation for Goal Attainment scores in the first and second interviews was .595 (N=89). When the discipline of the interviewer was held constant, the correlation was .629 for MSW interviewers (N=13) and .750 for RN interviewers (N=10). When the discipline of the interviewer in follow-up one was diffefent than that of the interviewer in follow-up two, the correlations were slightly lower.

b. The Combination Validity/Reliability Study mentioned previously gives another viewpoint on Reliability (Baxter, 1973). Here, therapist and the follow-up interviewers scored the Goal Attainment Follow-up Guides independently. Again, this is an extremely severe test of reliability, since there was a time span between the two interviews, and the sources of information were clearly much different for therapist and follow-up interviewers. Nevertheless, as Table VII illustrates, the correlations ranged from .46 to .85.

TABLE VII

Pearson Product-Moment Correlation between Goal Attainment Scores for clients based on independent (1) therapist-scorings and (2) follow-up interviewer scorings.

Adult OPD	N = 53	.507
Day Treatment	N = 8	.848
All Clients	N = 61	.458

These various results suggest a fair degree of reliability with internal consistency coefficients in the .60 range, and inter-rater agreement reliability coefficients in the .45 to .75 range. There seems to be sufficient Goal Attainment score consistency to make the methodology useful in a number of situations.

4. Appropriate Norms of Examinee Performance

Ebel's fourth subpoint under his general area of "The Meaningfulness of the Test Scores", is somewhat difficult to translate directly into Goal Attainment Scaling terms. Perhaps the most useful data on "norms" would be the mean Goal Attainment score of 50 and standard deviation of about 10, where each agency is seen as an "examinee". As mentioned earlier, most agencies using Goal Attainment Scaling have achieved such norms.

Another possible approach to establishing norms is some type of content analysis. A beginning on such norms is available from "Expectations and Goals for Clients at a Community Mental Health Service". (Garwick and Lampman, 1972) This study shows some norms for quantifiable variables for the Mental Health Service, but does not examine norms from other agencies, although Goal Attainment Follow-up Guides from several other agencies have been collected, and will be further analyzed in the future.

C. Technical Refinements, or the "Convenience of the Test"

Ebel's third major area of "test quality" re-

fers to the ease of the test's implementation and interpretation. Some improvements have been made in the convenience of Goal Attainment Scaling, most recently, Baxter's Conversion Key for Calculating Goal Attainment Scores from Unweighted Scores (Baxter (1973) and Garwick and Brintnall's Goal Attainment Scaling Calculation Tables (1973). In general, however, the convenience of the methodology could be improved further. The Goal Attainment Scaling methodology seems to be inherently attractive to many clinicians and administrators, but both (1) construction of the Goal Attainment Follow-up Guide and (2) interpretation of the Goal Attainment score may appear inconvenient to some.

1. Convenience of Construction of the Goal Attainment Follow-Up Guide

Some persons have praised the Goal Attainment Scaling concept because it enables evaluation to become part of the therapy and can be incorporated into the interaction with the client. Others, however, have complained of the difficulty of training personnel in Goal Attainment Scaling and in finding enough time to produce the Goal Attainment Follow-up Guides. One possible amelioration of this difficulty is the client construction of the follow-up guides, as illustrated by the Guide to Goals, Format One approach (Garwick, 1972).

The "review" or "monitoring" of the Goal Attainment Follow-up Guide is a closely related and difficult problem. As the Program Evaluation Project staff began accumulating Goal Attainment Follow-up Guides, it seemed that some follow-up guides produced by the clinicians included clerical or logical or descriptive shortcomings which rendered follow-up scoring very difficult or questionable ("Manual on Follow-up Assessment", Garwick, et. al. 1972). Various monitoring and follow-up guide revision techniques have been utilized, but none has been completely satisfactory, due to clinician dislike of the monitoring and a limited empirical basis of the monitoring criteria. The monitoring is expensive for the evaluation staff, and evidently unattractive to the clinical staff, yet the inconvenience and cost of follow-up guides with severe errors can be troublesome. Thus, review and revision of the Goal Attainment Follow-up Guide and the amount of time required for training and completing the follow-up guide are two of the more frequent complaints.

The client-specific nature of the Goal Attainment Scaling methodology, however, is apparently popular. One possibility for increasing clients involvement is the "programmed" Guide to Goals, Format One which was mentioned previously. If this device makes it possible to allow clients to construct their own follow-up guides with minimal clinical supervision, the convenience of Goal Attainment Scaling may be greatly increased. One possibility for minimizing the need for review is a "Semi-standardized" system where idiosyncratic (i.e., client-specific) construction is retained but a catalogue of possible Goal Attainment variables is used to decrease the the time required for follow-up guide construction.

2. Convenience of the Interpretation of the Goal Attainment Score

The Goal Attainment score is based on the formula from Kiresuk and Sherman's 1968 article. This formula seems formidable to some. One of the earliest attempts to ease Goal Attainment score calculation was the release of a short, simplified description of the calculation in simple steps. As mentioned above, Baxter has produced a conversion key and recently Sherman has suggested the possibility of simpler, scale-by-scale calculation of follow-up level score means for each follow-up guide. (See Chapter One of the P.E.P. Report, 1969-1973.) A series of tables which have been developed, make it possible to obtain the Goal Attainment score without calculation for Goal Attainment Follow-up Guides with from one to five scales (Garwick and Brintnall, 1973).

Because of the newness of Goal Attainment Scaling, comments on interpreting the Goal Attainment score have been purposely restricted. In general, the Goal Attainment score is said to be "the degree to which expectations for outcome at some certain time are reached".

Conclusion

Too often, it is forgotten that the Goal Attainment score is merely a tool, like the MMPI, the Strong Vocational Interest Blank, or the Wechsler Adult Intelligence Scale. Like any measurement tool, the Goal Attainment score is as useful as the design or system with which it is utilized. The degree of scientific irreproachability or clinical completeness depends on the way the methodology is integrated into a practical or experimental procedure.

References.

Anastasi, A. Psychological testing. (5th ed,) USA: Macmillan Company, 1970, P. 121.

Baxter, J. Goal attainment scaling conversion key for equally weighted scales. Unpublished Program Evaluation Project Report, February, 1973.

_____ Unpublished data report from the Program Evaluation Project, 1973.

_____ & Jones, S. Unpublished data report from the Program Evaluation Project, 1973.

_____ & Tripp, R. Unpublished data report from the Program Evaluation Project, 1972.

Campbell, D.T. & Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Cattel, R.B. Validity and reliability: a proposed more basic set of concepts. Journal of Educational Psychology, 1964, 55, 1-22.

Dreyer, J. Unpublished data report from the Program Evaluation Project, 1974.

Ebel, R.L. Must all tests be valid. American Psychologist, 1961, 16, 640-671.

Garwick, G. Guide to goals, format one. Unpublished Program Evaluation Project Report, October, 1972.

_____ Programmed instruction in goal attainment scaling. Unpublished Program Evaluation Project Report, October, 1973.

_____ & Brintnall, J. Tables for calculating the goal attainment score. Unpublished Program Evaluation Project Report, August, 1973.

_____, Grygelko, G., Makela, W., & Jones, S. Manual for the standardized assessment of the goal attainment follow-up guide. Unpublished Program Evaluation Project Report, July, 1972.

_____ & Lampman, S. Expectations and goals for clients at a community mental health service. Unpublished Program Evaluation Project Report, July, 1972.

Kiresuk, T.J. & Sherman, R.E. Goal attainment scaling: a general method for evaluating comprehensive community mental health programs. Community Mental Health Journal, 1968, 4(6), 443-453.

Lindquist, E.F. A first course in statistics (Rev. ed.) Boston: Houghton Mifflin, 1942.

Loevinger, J. Objective tests as instruments of psychological theory. Psychological Reports, 1957 Vol. 3, Monograph Supplement 9, 635-694.

Mauger, P. A study of the construct validity of Goal Attainment Scaling. Chapter Six, P.E.P. Report 1969-1973, June, 1974.

Mosier, C.I. A critical examination of the concepts of face validity. Educational & Psychological Measurement, 1947, 7, 191-205.

Nunnally, J.C. Psychometric theory. New York: McGraw-Hill, 1967. Pp. 78-79, 92.

Twedt, R. Unpublished data report from the Program Evaluation Project, 1974.

Wiggins, J.S. Personality & prediction, principles of personality assessment. USA: Addison-Wesley Company, 1973. Pp. 407-409.

PROGRAM EVALUATION PROJECT STAFF LISTING

CURRENT STAFF MEMBERS

Thomas J. Kiresuk, Ph.D.
Principal Investigator 1969-1974

Donna M. Audette
Research Assistant 1970
Follow-up Assistant 1971
Follow-up Supervisor 1972-1973
Utilization Consultant 1974

James Baxter
Research Assistant 1970
Operations Manager 1971
Assistant Coordinator 1972-1973
Re-design Coordinator 1974

Diane Berg
Editorial Secretary 1973-1974

Janis Bibee, M.A.
Editorial Secretary 1972
Editorial Assistant 1973
Assistant to the Editor 1974

Joan Brintnall
Student Assistant 1971
Secretary-Receptionist 1972
Administrative Assistant 1973
Dissemination, Consultation,
and Utilization Supervisor 1974

Joan Dreyer
Research Clerk 1972
Research Assistant 1973
Research/Administrative Clerk 1974

Geoffrey Garwick, M.A.
Research Applications Consultant 1970
Program Evaluation Coordinator 1971
Deputy Assistant Director 1972
Assistant Director 1973
Dissemination, Consultation,
and Utilization Consultant 1974

Carolyn Jasperson
Student Assistant 1973-1974

Susan Jones
Research Assistant 1971
Research Associate 1972-1974

Laurence Kivens, M.A.
Applications Analyst 1971
Editorial Supervisor 1971
Editor 1972-1974

Mary Knepper
Appointment Interviewer 1970-1973
Administrative Assistant 1974

Judy Long
Administrative Secretary 1974

Sander Lund
Management Applications Supervisor 1971
Assistant Coordinator 1971-1972
Coordinator for Administration 1973
Assistant Director 1974

Nancy Petersen
Secretary 1973
Research Assistant 1974

Michael Saunders
Programmer Analyst 1971-1972
Programmer Supervisor 1973-1974

Robert Sherman, Ph.D.
Associate Investigator 1969-1974

Roger Twedt
Student Assistant 1974

Mary Ellen Whalen
Student Assistant 1973-1974

PREVIOUS STAFF MEMBERS

Anita Bjornson
Research Assistant 1970

Barbara Blazick
Student Assistant 1970

Mary Duroche
Editorial Secretary 1972

David Feigal
Research Assistant 1970-1971

Thomas Griffin
Student Assistant 1970

Marilee Grygelko
Student Assistant 1971-1972
Research Assistant 1973

Edward Gubman
Student Assistant 1971

Colleen Halley
Student Assistant 1971-1973

Robert Kearney
Editorial Assistant 1972

Karen Kohout, M.A.
Research Analyst 1969
Research Supervisor 1970-1971

Sherry Lampman
Administrative Assistant 1970
Linguistic Analysis Consultant 1971
Content Analysis Supervisor 1972

William Makela, M.A.
Follow-up Supervisor 1970-1972

Charles Meade
Research Assistant 1970-1972

Deirdre Meade
Secretary-Receptionist 1971
Administrative Assistant 1972

Sylvia Muilenberg
Administrative Assistant 1969
Administrative Supervisor 1970

Nils Olsson
Student Assistant 1971
Research Assistant 1972-1973

Carol Pollock
Research Clerk 1972

William Prock
Community Applications Supervisor 1971-1972

Peter Ree
Student Assistant 1970

Martha Rose
Secretary-Receptionist 1970
Editorial Assistant 1971-1972

Susan Salasin
Coordinator 1969-1970
Assistant Director 1971-1972
Research Applications Consultant 1973

Richard Tripp
Programming Supervisor 1970
Design and Analysis Supervisor 1971-1972

Mary Trone
Editorial Secretary 1973

Carol Vanderpool
Secretary-Receptionist 1971-1972
Administrative Assistant 1973

Cynthia Wetterland
Secretary-Receptionist 1970

Allen Wichelman
Medical Records Clerk 1970
Management Applications Supervisor 1971

Sue Wright
Clinical Applications 1971

Carole Zimbrolt
Research Analyst 1969-1971

FOLLOW-UP INTERVIEWERS - CURRENT

Kathleen Bergum, M.S.W.

Charles Besnett, M.S.W.

Carol Dethmers, B.A.

Marcia Frankenberg, B.A.

George Meirick, M.S.W.

FOLLOW-UP INTERVIEWERS - PREVIOUS

Mary Ann Anzelc, R.N.

James Bergum, M.S.W.

Roanne Borkon, R.N.

Larry Bultena, M.S.W.

Scott Craven

Jeanine Emmons, R.N.

Barbara Gusek, R.N.

Mary Keturakat, R.N.

Steve Lapinsky

Betty Metz, B.A.

Madiline Sachs, R.N.

James Snope, M.S.W.

Milt Somerfleck, M.S.W.

CONSULTANTS

Dean Beaulieu, Ph.D.
Clinical Coordinator 1971-1974

James Boen, Ph.D.
Statistical Consultant 1969-1973

Byron Brown, Ph.D.
Statistical Consultant 1969-1973

Arthur Funke, Ph.D.
Dissemination Consultant 1969-1974

Stephen Greenwald, M.D.
Medications Consultant 1969-1973

Ann Russell
Clinical Consultant 1970

Robert Spano, A.C.S.W.
Patient Follow-up Coordinator 1969-1973

Wyman Spano
Editorial Consultant 1972-1973

Robert Walker, M.A.
Data Applications Coordinator 1973